JSC-08960

JSC INTERNAL NOTE NO. 74-FM-21

May 13, 1974

(NASA-TM-X-72176) A NEW ADAPTIVE CLASSIFIER USING ITERATIVE FILTERING (NASA) 25 p HC $3.25 CSCL 05B

N75-15324

Unclas

G3/60 06630

A NEW ADAPTIVE CLASSIFIER USING ITERATIVE FILTERING

Mathematical Physics Branch

MISSION PLANNING AND ANALYSIS DIVISION

National Aeronautics and Space Administration

LYNDON B. JOHNSON SPACE CENTER

Houston, Texas

JSC-08960

JSC INTERNAL NOTE NO. 74-FM-21

A NEW ADAPTIVE CLASSIFIER USING ITERATIVE FILTERING

By Arland L. Actkinson
Mathematical Physics Branch

May 13, 1974

MISSION PLANNING AND ANALYSIS DIVISION
NATIONAL AERONAUTICS AND SPACE ADMINISTRATION
JOHNSON SPACE CENTER
HOUSTON, TEXAS

Approved: ______________________
Emil R. Schiesser, Chief
Mathematical Physics Branch

Approved: J. R. Gurley ______________________
for Ronald L. Berry, Acting Chief
Mission Planning and Analysis Division

CONTENTS

A NEW ADAPTIVE CLASSIFIER USING ITERATIVE FILTERING

By Arland L. Actkinson

1.0 SUMMARY

To cope with signature variability, an algorithm has been defined which will adaptively classify remotely sensed data in the visible and near-infrared band. The signal is divided into a space-dependent component and a target-dependent component. The target-dependent component is assumed fixed across the image for each target type. The space-dependent component is estimated iteratively by a weighted, least-squares algorithm. Included in this study are the derivations of the sensor model and the two-dimensional, estimation algorithm.

2.0 INTRODUCTION: THE PROBLEM

The classification of remotely sensed image data, using current techniques, is severely hindered by the problem of signal variability. Signal variability refers to the vast differences in signals radiated by a single crop in response to different environmental conditions. Differences in soil type, local temperature, water content of target, amount of haze and cloud cover, and many other factors can have a decisive effect on the spectral response. In addition, changes in the angle at which the target is observed affect the signal.

One suggestion for dealing with signal variability is to use an adaptive classifier. An adaptive classifier alters the classification signatures in the course of an analysis. This usually means that after a resolution element is classified, the signature of the class is altered by averaging in the new observation. The averaging-in process may be weighted or unweighted. By using this technique, variations in signal can be partially modeled.

However, this adaptive classifier method has some drawbacks.

a. Errors in classification may cause a class signature to be "captured." For example, as Class A elements are misclassified as Class B, Class B statistics begin to look more and more like Class A, making future misclassification even more likely.

b. The weighting in the averaging is arbitrary; no criteria exist for this assignment, and, presumably, the values must be determined by experience.

c. This method presupposes very good initial estimates of the signatures.

d. This method restricts the taking of all ground truth to one part of the image, namely, the first part to be classified.

e. Few techniques account for variations due to angle of observation.

f. Observations of one class yield no information about the variations of signals from other classes. Consequently, if a class is not homogeneously distributed throughout an image, the signature statistics may not realistically reflect the environmental effects.

As a result of these disadvantages, a new model would seem desirable.

3.0 ANALYSIS[a]

Many of the difficulties of current classification methods are the results of trying to estimate the current signature of each class. If the class signatures could be considered as functions of some vector representing the environment, then signature modeling could be handled with statistical filters similarly to the measurement modeling performed in Apollo navigation. This should eliminate the previously mentioned problems of using the averaging technique.

A problem remains in implementing an environment estimation technique; namely, in order to model the signal, the class of the target being observed must be known. This means that, if n classes of targets are observed, then n signal types are to be processed in estimating the environment vector.

An analogous situation in navigation would be the following: Suppose several different types of measurements were being made (range, range rate, angles, etc.), but for each measurement, only the value was known, not the type. Before the observation could be used to update position and velocity, the measurement type must be identified. This would be done by asking which type of measurement was most reasonable, given the current estimate of position and velocity. The observation would be assumed to be this most reasonable type and would then be incorporated. This same procedure could be used in processing remotely sensed data. Thus,

1. First, determine what a measurement of each type (example, each crop classification) would look like, given the current estimate of the environment; in other words, estimate the signature of each class. If maximum likelihood is the classification criterion, then the mean and covariance matrix for each class should be determined.

2. Then, classify the measurement by using whatever classification criterion has been decided on.

3. Finally, incorporate the observation to refine the estimate of the environment state. The equations for this procedure are given in the following sections. The classification criterion is maximum likelihood.

[a]This analysis is given in reference 1.

4.0 SENSOR MODEL[a]

The model of a signal for visible and near visible light reflected from a target can be defined as

$$S = LRT \tag{1}$$

where

S is the signal received

L is the irradiance incident on the target

R is the reflectance of the target

T is the transmittance of the atmosphere from the target to the sensor

This model does not consider effects of the type where incoming radiation is absorbed by the target and radiated back at a different frequency. Visible light is seldom radiated in this manner. Path radiance is also not considered. Path radiance is the signal received by the sensor which is not from the target; for example, radiation from other points on the ground or from the sun, which, because of the aerosols in the atmosphere, is reflected into the sensor aperture. Although this effect is not as significant as the terms in equation (1), it can be important, and it is hoped that this effect will be added to the model at some future date.

Reassociating the terms in equation (1) gives

$$S = R(LT)$$

or

$$Y = R' + C$$

where

$Y = \ln S$

$R' = \ln R$

$C = \ln (LT)$

The signal Y, as well as R' and C, is a vector. Assume that the ith component of C has the form

$$C_i = a_i + b_iU + c_iV + d_iU^2 + e_iUV + f_iV^2 \tag{2}$$

where U and V are the spatial coordinates of the target point.

[a]This sensor model is taken from reference 2.

Hence,

$$Y = \begin{bmatrix} (R_1' + a_1), & b_1, & c_1, & d_1, & e_1, & f_1 \\ \vdots & & & & & \vdots \\ (R_n' + a_n), & b_n, & c_n, & d_n, & e_n, & f_n \end{bmatrix} \begin{bmatrix} 1 \\ U \\ V \\ U^2 \\ UV \\ V^2 \end{bmatrix} \tag{3}$$

The R' and a vectors are combined in equation (3) since neither is dependent on the position of the target. This gives the final form of the signal model, namely,

$$Y = \rho + \theta \tag{4}$$

where

Y is the signal received

$\rho = R' + a$

$\theta_i = b_i U + c_i V + d_i U^2 + e_i UV + f_i V^2$

remembering the assumption that

$$\ln (LT)_i = a_i + b_i U + c_i V + d_i U^2 + e_i UV + f_i V^2$$

Equation (4) will be the form used in this algorithm. Note that ρ is spatially independent; the difference between two ρ values is a function of target identity alone. On the other hand, θ is spatially dependent and target independent.

5.0 ESTIMATION ALGORITHM[a]

A digital image can be thought of as a matrix of numbers. For example, if the i, jth position is to be represented as dark, the number at that position would be small. The values of the matrix should be thougnt of as observations of a signal. If the signal is a linear function, or can be approximated as a linear function, the terms of the function can be estimated by least squares. If the image signal values are represented by the vector Y, then to say that the signal is a linear function of the vector G means there exists a matrix X such that

$$Y = XG \tag{5}$$

[a]This derivation was taken from reference 3.

If X is known, then the best estimate of G, in the least squares sense, is given by

$$\hat{G} = (X^T X)^{-1} X^T Y \qquad (6)$$

A sequential estimator seems an obvious choice; however, the choice of the sequence in which observations are processed is not obvious. The following ideas were used in selecting the sequence:

1. The data set used to generate G must be well defined.

2. The estimate of G and the computations used at an image point should be used in estimating the next image points.

3. Because modeling errors are always present, and because these errors probably increase as the distance from the estimation point increases, large jumps in the sequence should be avoided.

For these reasons, these choices were made:

1. The estimate of G at the i,jth point will be made by using the observations to the left and above that point; that is, over the set

$$\{(x,y) \mid 1 \le x \le i,\ 1 \le y \le j\} \qquad (7)$$

2. The estimate of G at $(i-1, j)$ and at $(i, j-1)$ will be used to find $\hat{G}$ at (i, j). It turns out the estimate at $(i-1, j-1)$ is also required.

5.1 Unweighted Estimation Algorithm

The set in expression (7) can be partitioned into

$A = \{(x,y) \mid 1 \le x < i,\ 1 \le y < j\}$

$B = \{(x,y) \mid 1 \le x < i,\ y = j\}$

$C = \{(x,y) \mid x = u,\ 1 \le y < j\}$

$D = \{(i,j)\}$

Clearly, the union of A, B, C, and D is equal to the set in expression (7), and no point in one of the partition sets is in any of the other sets.

For each partition, there exists a set of equations of the form of equation (1); the number of equations in the set is equal to the number of points in the partition. Hence,

$$
\begin{aligned}
Y_A &= X_A G \\
Y_B &= X_B G \\
Y_C &= X_C G \\
Y_D &= X_D G
\end{aligned}
\tag{8}
$$

where the elements in one of the Y_ℓ are the signals in set $\ell = A, B, C$, or D and the rows of the matrix X_ℓ are the coefficients of the elements of G. Recalling equation (6), the best estimate of G, using all of the observations in sets A, B, C, and D, is

$$
\hat{G} = \left\{ \begin{bmatrix} X_A \\ X_B \\ X_C \\ X_D \end{bmatrix}^T \begin{bmatrix} X_A \\ X_B \\ X_C \\ X_D \end{bmatrix} \right\}^{-1} \begin{bmatrix} X_A \\ X_B \\ X_C \\ X_D \end{bmatrix}^T \begin{bmatrix} Y_A \\ Y_B \\ Y_C \\ Y_D \end{bmatrix}
$$

$$
\hat{G} = \left[X_A^T X_A + X_B^T X_B + X_C^T X_C + X_D^T X_D \right]^{-1} \left[X_A^T Y_A + X_B^T Y_B + X_C^T Y_C + X_D^T Y_D \right] \tag{9}
$$

Both of the terms in equation (9) use quantities which will be inconvenient to define, namely, X_B, X_C, Y_A, Y_B, and Y_C. Fortunately, equation (9) can be rewritten. Using X_{AB} to stand for X in equation (9) over the set $A \cup B$,

$$
X_{AB}^T X_{AB} = X_A^T X_A + X_B^T X_B
$$

Similarly,

$$
X_{AC}^T X_{AC} = X_A^T X_A + X_C^T X_C
$$

Likewise, $\hat{G}_{AB}$, $\hat{G}_{AC}$, and $\hat{G}_A$ will denote the best estimates of G, using the observations in $A \cup B$, $A \cup C$, and A, respectively. Hence, the inverse term in equation (9) becomes

$$\left[X_A^T X_A + X_B^T X_B + X_C^T X_C + X_D^T X_D\right]^{-1} = \left[(X_A^T X_A + X_B^T X_B) + (X_A^T X_A + X_C^T X_C) - X_A^T X_A + X_D^T X_D\right]^{-1}$$

$$= \left[X_{AB}^T X_{AB} + X_{AC}^T X_{AC} - X_A^T X_A + X_D^T X_D\right]^{-1}$$

The second factor in equation (9) also needs to be rewritten. Recall that

$$X_A^T X_A \hat{G}_A = X_A^T Y_A$$

$$X_{AB}^T X_{AB} \hat{G}_{AB} = X_{AB}^T Y_{AB} = X_A^T Y_A + X_B^T Y_B$$

$$X_{AC}^T X_{AC} \hat{G}_{AC} = X_{AC}^T Y_{AC} = X_A^T Y_A + X_C^T Y_C$$

Hence, the second term becomes

$$X_A^T Y_A + X_B^T Y_B + X_C^T Y_C + X_D^T Y_D = (X_A^T Y_A + X_B^T Y_B) + (X_A^T Y_A + X_C^T Y_C) - X_A^T Y_A + X_D^T Y_D$$

$$= X_{AB}^T X_{AB} \hat{G}_{AB} + X_{AC}^T X_{AC} \hat{G}_{AC} - X_A^T X_A \hat{G}_A + X_D^T Y_D$$

Substituting these rewritten terms into equation (9) give

$$G = \left[X_{AB}^T X_{AB} + X_{AC}^T X_{AC} - X_A^T X_A + X_D^T X_D\right]^{-1} \left[X_{AB}^T X_{AB} \hat{G}_{AB} + X_{AC}^T X_{AC} \hat{G}_{AC} - X_A^T X_A \hat{G}_A + X_D^T Y_D\right] \tag{10}$$

Equation (10) is the form that will be used in determining $\hat{G}$, assuming $Y = XG$, where X is some matrix and the equations in the system are to be equally weighted.

5.2 Weighted Estimation Algorithm

In equation (10) each observation has exactly the same importance in calculating $\hat{G}$. This is reasonable if all the observations are equally good. However, more distant pixels would be expected to be less useful than closer pixels. Consequently, distant measurements should be downweighted.

Suppose G is to be estimated at some point $P_n = (U_n, V_n)$ in the image. Select a number r such that $0 < r < 1$. If $d(i,n)$ is the distance, by whatever definition, between image points P_i and P_n, then the observation at P_i may be weighted by $r^{d(i,n)}$. Using this weighting, the greater the distance of an observation from P_n, the smaller the weight, until finally the very distant observations are, effectively, thrown away.

The remaining task is to define the distance measure $d(i,n)$. Suppose $P_1 = (U_1,V_1)$ and $P_2 = (U_2,V_2)$ are two points in the image. Define

$$d(1,2) = |U_1 - U_2| + |V_1 - V_2| \tag{11}$$

The convenience of this distance measure will become apparent.

Equation (5), at point $P_n = (U_n,V_n)$, now becomes

$$W(U_n,V_n)Y = W(U_n,V_n)XG \tag{12}$$

where

$$W(U_n,V_n) = \begin{bmatrix} r^{d(n,1)} & & 0 \\ & \ddots & \\ 0 & & r^{d(n,n)} \end{bmatrix} \tag{13}$$

Here $W(U_n,V_n)$ is the diagonal matrix with the general term $r^{d(n,i)}$.

The best estimate of G in equation (12) is

$$\hat{G} = [X^T W(U_n,V_n)^2 X]^{-1} X^T W(U_n,V_n)^2 Y$$

Notice that each diagonal term of $W(U,V)$ changes as the G estimation point changes. Suppose G is estimated at P_n. Then the ith diagonal term of $W(U_n,V_n)$ is $r^{d(i,n)}$. In other words, $r^{d(i,n)}$ is the weight of the ith observation, when estimating G at point P_n. Suppose G is now to be estimated at another point Q which is a distance of $d(i,n) + 1$ from P_i. The the appropriate weight for the ith observation in computing G at point Q is $r^{d(i,n)+1}$. Hence, the ith diagonal element of the weighting matrix at Q is r times the ith diagonal element of the weighting matrix at P_n.

Suppose G is to be estimated at (U,V). Further, suppose G has been estimated at $(U,V-1)$, $(U-1,V)$, and $(U-1,V-1)$, which means that weighting matrices have been defined at these points also. Use the partitioning described in the unweighted case,

$A = \{(x,y) \mid 1 \le x < U,\ 1 \le y < V\}$

$B = \{(x,y) \mid 1 \le x < U,\ y = V\}$

$C = \{(x,y) \mid X = u,\ 1 \le y < V\}$

$D = \{(U,V)\}$

The estimates of G at $(U-1, V-1)$, $(U-1, V)$, and $(U, V-1)$, respectively, are

$$\hat{G} \text{ at } (U-1, V-1) = \left[X_A^T W_{(U-1,\ V-1)}^2 X_A\right]^{-1} X_A^T W_{(U-1,\ V-1)}^2 Y_A$$

$$\hat{G} \text{ at } (U-1, V) = \left[X_{AB}^T W_{(U-1,\ V)}^2 X_{AB}\right]^{-1} X_{AB}^T W_{(U-1,V)}^2 Y_{AB}$$

$$\hat{G} \text{ at } (U, V-1) = \left[X_{AC}^T W_{(U,\ V-1)}^2 X_{AC}\right]^{-1} X_{AC}^T W_{(U,\ V-1)}^2 Y_{AC} \qquad (14)$$

where Y_A, Y_{AB}, and Y_{AC} are the observations from the sets A, $A \cup B$, and $A \cup C$, respectively.

Denote by R the matrix

$$R = \begin{bmatrix} r & & & 0 \\ & r & & \\ & & \ddots & \\ 0 & & & r \end{bmatrix}$$

where r is the weighting scalar mentioned previously. The size of R will be implicitly defined by the equation in which it appears.

Recall that $W_{(U-1,\ V-1)}$, $W_{(U-1,\ V)}$, and $W_{(U,\ V-1)}$ have the weights for the elements in Y_A, Y_{AB}, and Y_{AC}, respectively. Partitioning $W_{(U-1,\ V)}$ and $W_{(U,\ V-1)}$ gives

$$W_{(U-1,\ V)} = \left[\begin{array}{c|c} W_{(U-1,\ V)}(A) & 0 \\ \hline 0 & W_{(U-1,\ V)}(B) \end{array}\right]$$

and

$$W_{(U,\ V-1)} = \left[\begin{array}{c|c} W_{(U,\ V-1)}(A) & 0 \\ \hline 0 & W_{(U,\ V-1)}(C) \end{array}\right]$$

where $W_{(U-1,\ V)}(A)$ and $W_{(U-1,\ V)}(B)$ are the diagonal weighting matrices for Y_A and Y_B, respectively, defined at $(U-1, V)$, and $W_{(U,\ V-1)}(A)$ and $W_{(U,\ V-1)}(C)$ are the diagonal weighting matrices for Y_A and Y_C, respectively, defined at $(U, V-1)$.

Note that the distance from $(U-1, V-1)$ to (U, V) is 2. Furthermore, the distance from (U, V) to any point in A is 2 greater than the distance $d(i, n)$ from $(U-1, V-1)$ to that point in A. Hence, the weight for an observation at any point P_i in A is $r^2r^{d(i,n)}$. Notice that $r^{d(i,\ n)}$ is the general diagonal term of $W_{(U-1,\ V-1)}(A) = W_{(U-1,\ V-1)}$. This means

$$W_{(U,\ V)}(A) = R^2W_{(U-1,\ V-1)}(A) \tag{15}$$

Similarly, because the distance from (U, V) to a point P_i in $A \cup B$ or $A \cup C$ is 1 greater than the distance from $(U-1, V)$ or $(U, V-1)$, respectively, to P_i,

$$\left.\begin{aligned} W_{(U,\ V)}(A) &= RW_{(U,\ V-1)}(A) = RW_{(U-1,\ V)}(A) \\ W_{(U,\ V)}(B) &= RW_{(U,\ V-1)}(B) \\ W_{(U,\ V)}(C) &= RW_{(U-1,\ V)}(C) \end{aligned}\right\} \tag{16}$$

Now the machinery is gathered for defining $\hat{G}$ at (U, V). Recall that the best estimate of G at (U, V) is

$$\hat{G} = \left\{ \left[X_A^T,\ X_B^T,\ X_C^T,\ X_D^T\right] \begin{bmatrix} W^2_{(U,\ V)}(A) & & & 0 \\ & W^2_{(U,\ V)}(B) & & \\ & & W^2_{(U,\ V)}(C) & \\ 0 & & & W^2_{(U,\ V)}(D) \end{bmatrix} \begin{bmatrix} X_A \\ X_B \\ X_C \\ X_D \end{bmatrix} \right\}^{-1}$$

$$\left[X_A^T,\ X_B^T,\ X_C^T,\ X_D^T\right] \begin{bmatrix} W^2_{(U,\ V)}(A) & & & 0 \\ & W^2_{(U,\ V)}(B) & & \\ & & W^2_{(U,\ V)}(C) & \\ 0 & & & W^2_{(U,\ V)}(D) \end{bmatrix} \begin{bmatrix} Y_A \\ Y_B \\ Y_C \\ Y_D \end{bmatrix}$$

$$= \left[X_A^T W^2_{(U,\ V)}(A) X_A + X_B^T W^2_{(U,\ V)}(B) X_B + X_C^T W^2_{(U,\ V)}(C) X_C \right.$$

$$\left. + X_D^T W^2_{(U,\ V)}(D) X_D \right]^{-1} \left[X_A^T W^2_{(U,\ V)}(A) Y_A + X_B^T W^2_{(U,\ V)}(B) Y_B \right.$$

$$\left. + X_C^T W^2_{(U,\ V)}(C) Y_C + X_D^T W^2_{(U,\ V)}(D) Y_D \right] \qquad (17)$$

Recall that

$$X_{AB}^T W^2_{(U,\ V)}(A \cup B) X_{AB} = X_A^T W_{(U,\ V)}(A) X_A + X_B^T W_{(U,\ V)}(B) X_B$$

$$= X_A^T [R^2 W^2_{(U-1,\ V)}(A)] X_A + X_B^T [R^2 W^2_{(U-1,\ V)}(B)] X_B$$

$$= R^2 X_A^T W^2_{(U-1,\ V)}(A) X_A + R^2 X_B^T W^2_{(U-1,\ V)}(B) X_B$$

Similarly,

$$X_{AC}^T W^2_{(U,\ V)}(A \cup C) X_{AC} = X_A^T [R^2 W^2_{(U,\ V-1)}(A)] X_A + X_C^T [R^2 W^2_{(U,\ V-1)}(C)] X_C$$

$$= R^2 X_A^T W^2_{(U,\ V-1)}(A) X_A + R^2 X_C^T W^2_{(U,\ V-1)}(C) X_C$$

Also,

$$X_A^T W^2_{(U,\ V)}(A) X_A = X_A^T [R^4 W^2_{(U-1,\ V-1)}(A)] X_A$$

$$= R^4 X_A^T W^2_{(U-1,\ V-1)}(A) X_A$$

Now the inverse factor of equation (17) can be rewritten as

$$X_A^T W^2_{(U,\ V)}(A) X_A + X_B^T W^2_{(U,\ V)}(B) X_B + X_C^T W^2_{(U,\ V)}(C) X_C + X_D^T W^2_{(U,\ V)}(D) X_D$$

$$= X_A^T R^2 W^2_{(U-1,\ V)}(A) X_A + X_B^T R^2 W^2_{(U-1,\ V)}(B) X_B$$

$$+ X_A^T R^2 W^2_{(U,\ V-1)}(A) X_A + X_C^T R^2 W^2_{(U,\ V-1)}(C) X_C$$

$$- X_A^T R^2 W^2_{(U-1,\ V)}(A) X_A + X_D^T W^2_{(U,\ V)}(D) X_D$$

$$= R^2 \left[X_A^T W^2_{(U-1,\ V)}(A) X_A + X_B^T W^2_{(U-1,\ V)}(B) X_B \right]$$

$$+ R^2 \left[X_A^T W^2_{(U,\ V-1)}(A) X_A + X_C^T W^2_{(U,\ V-1)}(C) X_C \right]$$

$$- X_A^T R^4 W^2_{(U-1,\ V-1)}(A) X_A + X_D^T W^2_{(U,\ V)}(D) X_D$$

$$= R^2 \left[X_{AB}^T W^2_{(U-1,\ V)}(A \cup B) X_{AB} + X_{AC}^T W^2_{(U,\ V-1)}(A \cup C) X_{AC} \right]$$

$$- R^4 \left[X_A^T W^2_{(U-1,\ V-1)}(A) X_A \right] + X_D^T W^2_{(U,\ V)}(D) X_D \qquad (18)$$

Now, the first factor of equation (17) is in terms of matrices that were defined at (U, V-1), (U-1, V), or (U-1, V-1) plus the new term defined at (U, V). This is the form of the inverse term which will be used.

All that remains is to rewrite the second factor in equation (17). We know

$$X_A^T W^2_{(U-1,\ V-1)}(A) Y_A = \left[X_A^T W^2_{(U-1,\ V-1)}(A) X_A^T \right] \hat{G}_A$$

$$X_{AB}^T W^2_{(U-1,\ V)}(A \cup B) Y_{AB} = \left[X_{AB}^T W^2_{(U-1,\ V)}(A \cup B) X_{AB}^T \right] \hat{G}_{AB}$$

$$X_{AC}^T W^2_{(U,\ V-1)}(A \cup C) Y_{AC} = \left[X_{AC}^T W^2_{(U,\ V-1)}(A \cup C) X_{AC}^T \right] \hat{G}_{AC}$$

where $\hat{G}_A$, $\hat{G}_{AB}$, and $\hat{G}_{AC}$ are the best estimates of G over A, $A \cup B$, and $A \cup C$ at (U-1, V-1), (U, V-1), and (U-1, V), respectively.

Adjusting the first equation to change the weighting matrix reference point to (U, V),

$$R^4 X_A^T W^2_{(U-1,\ V-1)}(A) Y_A = R^4 \left[X_A^T W^2_{(U-1,\ V-1)}(A) X_A \right] \hat{G}_A$$

$$X_A^T \left[R^4 W^2_{(U-1,\ V-1)}(A) \right] Y_A = R^4 \left[X_A^T W^2_{(U-1,\ V-1)}(A) X_A \right] \hat{G}_A$$

$$X_A^T W^2_{(U,\ V)}(A) Y_A = R^4 \left[X_A^T W^2_{(U-1,\ V-1)}(A) X_A \right] \hat{G}_A$$

Similarly,

$$X_{AB}^T W^2_{(U,\ V)}(A \cup B) Y_{AB} = R^2 X_{AB}^T W^2_{(U-1,\ V)}(A \cup B) X_{AB} \hat{G}_{AB}$$

and

$$X_{AC}^T W^2_{(U,\ V)}(A \cup C) Y_{AC} = R^2 X_{AC}^T W^2_{(U,\ V-1)}(A \cup C) X_{AC} \hat{G}_{AC}$$

The second factor in (13) can now be rewritten

$$\begin{aligned}
& X_A^T W^2_{(U,\ V)}(A) Y_A + X_B^T W^2_{(U,\ V)}(B) Y_B + X_C^T W^2_{(U,\ V)}(C) Y_C + X_D^T W^2_{(U,\ V)}(D) Y_D \\
&= \left[X_A^T W^2_{(U,\ V)}(A) Y_A + X_B^T W^2_{(U,\ V)}(B) Y_B \right] \\
&\quad + \left[X_A^T W^2_{(U,\ V)}(A) Y_A + X_C^T W^2_{(U,\ V)}(C) Y_C \right] \\
&\quad - X_A^T W^2_{(U,\ V)}(A) Y_A + X_D^T W^2_{(U,\ V)}(D) Y_D \\
&= X_{AB}^T W^2_{(U,\ V)}(A \cup B) Y_{AB} \\
&\quad + X_{AC}^T W^2_{(U,\ V)}(A \cup C) Y_{AC} - X_A^T W^2_{(U,\ V)}(A) Y_A \\
&\quad + X_D^T W^2_{(U,\ V)}(D) Y_D \\
&= R^2 \left[X_{AB}^T W^2_{(U-1,\ V)}(A \cup B) X_{AB} \hat{G}_{AB} + X_{AC}^T W^2_{(U,\ V-1)}(A \cup C) X_{AC} \hat{G}_{AC} \right] \\
&\quad - R^4 \left[X_A^T W^2_{(U-1,\ V-1)}(A) X_A \hat{G}_A \right] + X_D^T W^2_{(U,\ V)}(D) Y_D
\end{aligned} \tag{19}$$

Substituting (18) and (19) into (17), the best estimate of G at (U, V), in the weighted least squares sense, is

$$\hat{G} = \left\{ R^2 \left[X_{AB}^T W^2_{(U-1,\, V)}(A \cup B) X_{AB} + X_{AC}^T W^2_{(U,\, V-1)}(A \cup C) X_{AC} \right] - R^4 \left[X_A^T W^2_{(U-1,\, V-1)}(A) X_A \right] + X_D^T W^2_{(U,\, V)}(D) X_D \right\}^{-1} \left\{ R^2 \left[X_{AB}^T W^2_{(U-1,\, V)}(A \cup B) X_{AB} \hat{G}_{AB} + X_{AC}^T W^2_{(U,\, V-1)}(A \cup C) X_{AC} \hat{G}_{AC} \right] - R^4 \left[X_A^T W^2_{(U-1,\, V-1)}(A) X_A \hat{G}_A \right] + X_D^T W^2_{(U,\, V)}(D) X_D \right\} \tag{20}$$

6.0 IMPLEMENTATION

To complete the definition of the adaptive classifier, the method of handling the training data will be defined. Finally, the procedure for classification described in section 3.0 will be described.

6.1 Training Data Calculations

First, the signatures of the classification sets must be estimated. This should be accomplished by evaluating the equations

$$\begin{bmatrix} \rho_1 \\ \cdot \\ \cdot \\ \cdot \\ \rho_n \\ b \\ c \\ d \\ e \\ f \end{bmatrix} = (H^T H)^{-1} H^T Y \tag{21}$$

where

ρ_i = the log of the reflectance of the ith class plus a

b, c, d, e, f = the coefficients of θ

Y = the vector of the observations

H = the matrix over the training data such that the jth line is

$$[I_j, U_j, V_j, U_j^2, U_jV_j, V_j^2]$$

where

I_j = a row vector with all zeroes except for a one in the ith location, indicating the jth observation came from the ith class

(U_j, V_j) = the position, in U,V coordinates, of the jth observation

Equations (5) and (6) show that equation (21) is a solution of

$$Y = H \begin{bmatrix} \rho_1 \\ \cdot \\ \cdot \\ \cdot \\ \rho_n \\ b \\ c \\ d \\ e \\ f \end{bmatrix}$$

The solution of equation (21) will yield an estimate of the class reflectances and the global estimate of the environment function θ. In order to use a maximum likelihood classifier, as described, for example, in reference 4, the covariance matrices of the classes are also needed. These may be estimated by

$$C_i = E_i\left[\left\{(Y - \theta) - \rho_i\right\}\left\{(Y - \theta) - \rho_i\right\}^T\right]$$

where

C_i = the covariance matrix of the ith class

$E_i[*]$ = the expected value of * taken over the ith class

Y = the log of the received signal

θ = the estimate of the environment; i.e., $bU + cV + dU^2 + eUV + fV^2$

ρ_i = the log of the average reflectance of the ith class plus a

Finally, the best local estimate of the environment vector must be made at point (1,1), since that is the point where classification will be. The best local estimate at (1,1) will be a solution of

$$W_iY_i = W_i[\rho_i,1,U_i,V_iU_i^2,U_iV_i,V_i^2]\begin{bmatrix}1\\ \Delta a\\ b\\ c\\ d\\ e\\ f\end{bmatrix} \qquad (22)$$

where

W_i = the weight of the ith observation (equal to $\beta^{(U_i + V_i - 2)}$, $0 < \beta \leq 1$)

Y_i = the ith observation (scalar) for the channel being considered

ρ_i = the logarithm of the reflectance of the object of the ith observation plus a

(U_i,V_i) = the coordinates of the ith observation

$[a,\ldots,f]^T$ = the vector of coefficients of the log environment polynomial; that i., $\theta = a + bU + cV + dU^2 + eUV + fV^2$

The solution of equation (22) is

$$\begin{bmatrix}\Delta a\\ b\\ c\\ d\\ e\\ f\end{bmatrix} = (K^TW^2K)^{-1}K^TW^2Y \qquad (23)$$

where

$$W = \begin{bmatrix}\ddots & & 0\\ & \beta^{U+V-2} & \\ 0 & & \ddots\end{bmatrix}$$

and

$$K = \begin{bmatrix} & \vdots & \\ \rho_i,1,U_i,V_i,U_i,U_iV_i,V_i^2 \\ & \vdots & \end{bmatrix}$$

A couple of comments are necessary here. First, in general, $r < \beta$, where r is defined in section 5.2. This means that in estimating the environment using training data, the old measurements are less downweighted than they would be in the normal, weighted estimation procedure. This reflects the greater security resulting from knowing the classification of the data. Second, for the local estimates of the environment, a constant term will be used in the polynomial. The globally defined polynomial also has a constant term, but the term is combined with the reflectance.

6.2 Test Data Calculations

The calculations to be performed on the data to be classified are

1. Estimate the environment state from prior observations.

2. Classify the observation.

3. Update the estimate of the environment.

The equations to perform these functions are as follows:

Suppose the observation at coordinates (U, V) is to be processed. Denote by $\hat{G}_{(U, V)}$ the best estimate of the environment state vector at (U, V). Then,

$$\hat{G}^{-}_{(U, V)} = \frac{\hat{G}_{(U-1, V)} + \hat{G}_{(U, V-1)}}{2} \tag{24}$$

The minus sign indicates that the observation at (U, V) has not yet been used in the estimate of G.

The estimate $\hat{G}^{-}_{(1,1)}$ is determined from the training data only. In the event that $U = 1$, then $\hat{G}^{-}_{(U, V)} = \hat{G}_{(U, V-1)}$. For the case where $V = 1$, then $\hat{G}^{-}_{(U, V)} = \hat{G}_{(U-1, V)}$.

Now that G has been estimated, θ can be calculated where

$$\theta + a = [1,U,V,U^2,UV,V^2]\hat{G}^{-}_{(U, V)} \tag{25}$$

Then the number ρ may be estimated by

$$\rho = Y - \theta \tag{26}$$

where Y is the log of the observed signal.

Finally, the likelihood of each class must be computed.

$$L(i|\rho) = -\ln(\Sigma_i) - (\rho - \rho_i)^T \Sigma_i^{-1} (\rho - \rho_i) \qquad (27)$$

where

$L(i|\rho)$ = the likelihood of the observation ρ belonging to the ith class

ρ_i = the mean of the ith class

Σ_i = the covariance matrix of the ith class

The observation should be classified into the class i such that

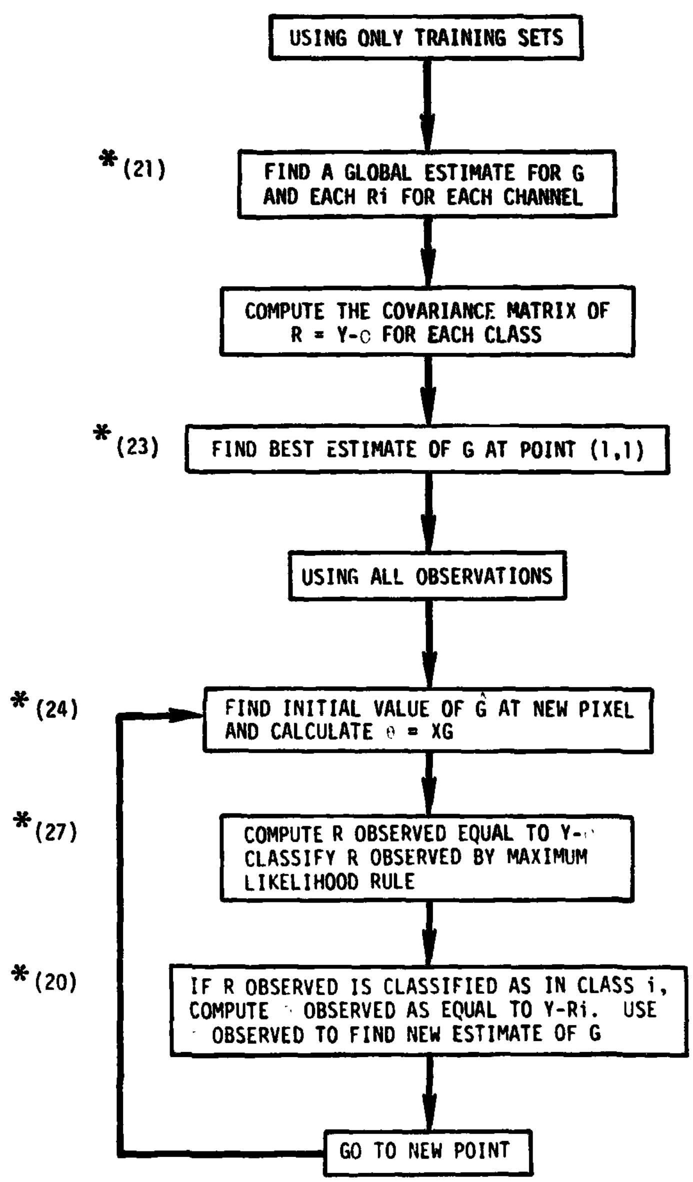
$$L(i|\rho) > L(j|\rho), \; i \neq j$$

In the case where $L(i|\rho) = L(j|\rho)$, some arbitrary choice between i and j must be made.

Now that the classification has been established, the observation may be used to update the estimate of G. This is done by means of equation (20). A descriptive flow chart of the estimation algorithm is given in figure 1. Details are given in figure 2.

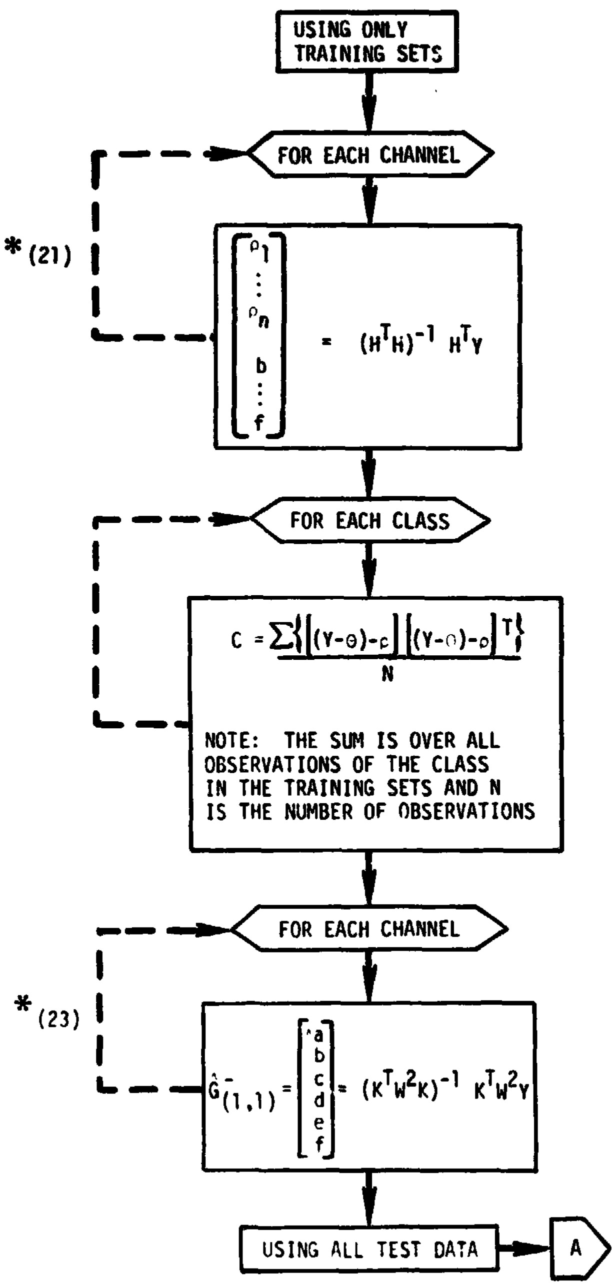
7.0 CONCLUSION

To cope with signature variability, an algorithm has been defined which will adaptively classify remotely sensed data in the visible and near-infrared band. The signal is divided into a space-dependent component and a target-dependent component. The target-dependent component is assumed fixed across the image for each target type. The space-dependent component is estimated iteratively by a weighted, least-squares algorithm. Included in the study was the derivation of the sensor model and two-dimensional, estimation algorithm.



* NUMBERS IN PARENTHESE REFER
TO EQUATIONS IN THE TEXT

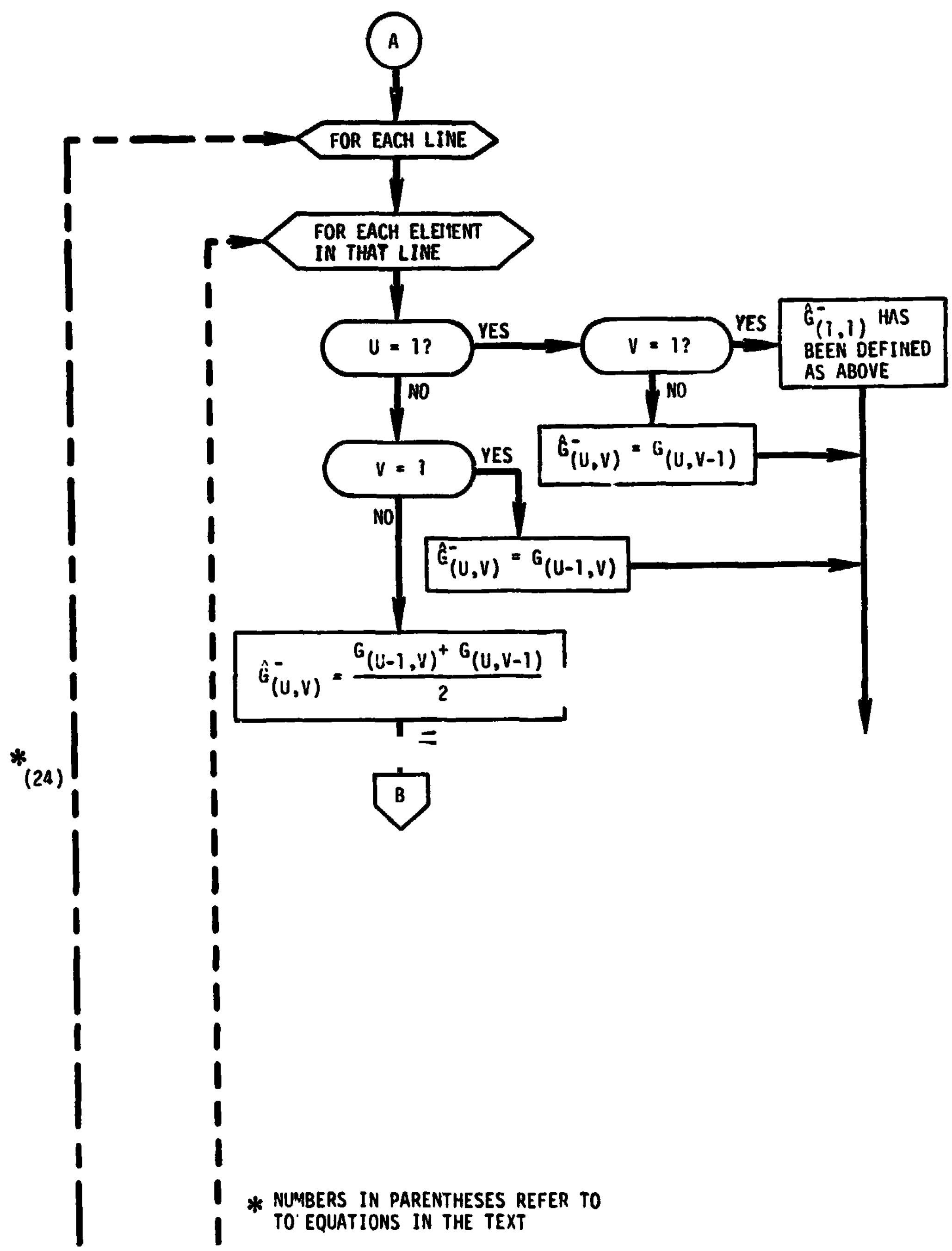
Figure 1.- Descriptive flow chart of estimation algorithm.



* NUMBERS IN PARENTHESES REFER TO EQUATIONS IN THE TEXT

Page 1 of 3

Figure 2.- Detailed flow chart of estimation algorithm.



* NUMBERS IN PARENTHESES REFER TO TO EQUATIONS IN THE TEXT

Figure 2.- Continued.

B

$$\Theta = \left|1,\ U,\ V,\ U^2,\ UV,\ V^2\ \ \hat{G}^{-}_{(U,V,)}\right|$$

$\rho = Y - \Theta$, FOR ALL CHANNELS

CLASSIFY BY FINDING THE CLASS i
SUCH THAT $L(i,\rho) > L(j,\rho),\ i \neq j$

WHERE

$$L(i,\rho) = -\ln\left|\Sigma i\right| - (\rho-\rho)^T \ \ \Sigma^{-1}(\rho-\rho)$$

$$\begin{aligned}\hat{G} = \Big\{ & R^2\left[X_{AB}{}^T W^2_{(U-1,\ V)}(AUB)X_{AB} + X_{AC}{}^T W^2_{(U,\ V-1)}(AUC)X_{AC}\right] \\ & - R^4\left[X_A{}^T W^2_{(U-1,\ V-1)}(A)X_A\right] + X_D{}^T W^2_{(U,\ V)}(D)X_D\Big\}^{-1} \\ \Big\{ & R^2\left[X_{AB}{}^T W^2_{(U-1,\ V)}(AUB)X_{AB}\hat{G}_{AB} + X_{AC}{}^T W^2_{(U,\ V-1)}(AUC)X_{AC}\hat{G}_{AC}\right] \\ & - R^4\left[X_A{}^T W^2_{(U-1,\ V-1)}(A)\ X_A\hat{G}_A\right] + X_D{}^T W^2_{(U,\ V)}(D)\ X_D\Big\}\end{aligned}$$

* (25)
* (26)
* (27)
* (20)

* NUMBERS IN PARENTHESES REFER TO EQUATIONS IN THE TEXT

Figure 2.- Concluded.

REFERENCES

1. Actkinson, Arland L.: A New Adaptive Classifier Using Kalman Filtering. JSC memo FM84(73-209), Aug. 20, 1973.

2. Actkinson, Arland L.: Sensor Model for Remotely Sensed Observations for Use in an Adaptive Classification Algorithm. JSC memo FM84(73-379), Ja [illegible], 1974.

3. Actkinson, Arland L.: Estimation Algorithm for Image Analysis. JSC memo FM84(73-380), Feb. 7, 1974.

4. Fu, K. S.; Landgrebe, D. A.; and Phillips, T. L.: Information Processing of Remotely Sensed Agricultural Data. Proceedings of the IEEE, Vol. 57, No. 4, Apr. 1969.